

## Analisis Potensi *Marketplace* Terhadap Penjualan Menggunakan *K-Nearest Neighbors*

Sya'roni<sup>1</sup>,

Institut Teknologi Dan Bisnis Mesuji

e-mail: <sup>1</sup>rsya1522@gmail.com

Diajukan: 19 Agustus 2025; Direvisi: 02 Oktober 2025; Diterima: 04 Oktober 2025

### Abstrak

Perkembangan teknologi informasi dan komunikasi telah mengubah pola konsumsi Masyarakat, terutama dalam berbelanja daring melalui platform *e-commerce* di Indonesia seperti Shopee, Tokopedia, dan Bukalapak. Namun, pelaku usaha masih kesulitan menentukan platform yang paling potensial untuk memasarkan produknya akibat kurangnya informasi dan kemampuan analisis data. Penelitian ini bertujuan untuk menganalisis potensi pasar produk di ketiga marketplace tersebut menggunakan algoritma *K-Nearest Neighbors* (KNN). Data yang digunakan diperoleh melalui scraping dari masing-masing platform pada tahun 2022, dengan total 2.250 record yang mencakup variabel harga, jumlah penjualan, ulasan, dan diskon. Hasil analisis berdasarkan marketplace dan kategori menunjukkan pola yang jelas, Shopee unggul pada kategori fashion (skor 0,5293), Bukalapak pada elektronik (skor 1,0000), dan Tokopedia pada otomotif (skor 0,1653). Evaluasi menunjukkan bahwa model KNN mampu mengklasifikasikan produk ke dalam kategori terlaris dan tidak terlaris dengan akurasi sebesar 84,62%, dan nilai AUC sebesar 0,9185, yang menunjukkan kemampuan model dalam membedakan produk dengan potensi penjualan tinggi sedang dan rendah. Temuan ini memberikan wawasan strategis bagi pelaku usaha dalam pengambilan keputusan terkait pemasaran dan manajemen stok. Studi ini menekankan pentingnya menggunakan analisis berbasis data untuk meningkatkan daya saing di pasar digital.

**Kata kunci:** KNN, marketplace, produk

### Abstract

The development of information and communication technology has changed people's consumption patterns, especially in online shopping through *e-commerce* platforms in Indonesia such as Shopee, Tokopedia, and Bukalapak. However, business actors still struggle to determine the most potential platforms to market their products due to a lack of information and data analysis capabilities. This study aims to analyze the market potential of products in these three marketplaces using the *K-Nearest Neighbors* (KNN) algorithm. The data used was obtained through scraping from each platform in 2022, with a total of 2,250 records covering the variables of price, number of sales, reviews, and discounts. The analysis results by marketplace and category show a clear pattern, with Shopee excelling in the fashion category (score 0.5293), Bukalapak in electronics (score 1.0000), and Tokopedia in automotive (score 0.1653). The evaluation shows that the KNN model is able to classify products into best-selling and non-best-selling categories with an accuracy of 84.62%, and an AUC value of 0.9185, indicating the model's ability to distinguish products with high, medium, and low sales potential. These findings provide strategic insights for businesses in making decisions related to marketing and inventory management. The study emphasizes the importance of using data-driven analytics to increase competitiveness in the digital marketplace.

**Keywords:** KNN, marketplace, product

### 1. Pendahuluan

Perubahan signifikan dalam pola konsumsi konsumen terutama dalam aktivitas pembelian, telah didorong oleh kemajuan teknologi informasi dan komunikasi. Pelanggan kini dapat bertransaksi dengan lebih mudah dan tanpa batasan waktu atau geografis berkat maraknya platform *e-commerce* di Indonesia. Marketplace yang menyediakan beragam produk, opsi pembayaran, dan promosi menarik untuk menarik

pelanggan, seperti Shopee, Tokopedia, dan Bukalapak, telah muncul sebagai pilar penting pertumbuhan ekonomi digital Indonesia.

Transformasi digital yang pesat telah mendorong pergeseran preferensi konsumen dari metode belanja konvensional ke pasar daring. Fenomena ini semakin nyata di Indonesia berkat meningkatnya akses internet, penetrasi ponsel pintar, dan kemudahan metode pembayaran digital. Selama sepuluh tahun terakhir, perkembangan teknologi informasi dan komunikasi telah membawa perubahan signifikan dalam berbagai aspek kehidupan masyarakat, termasuk pola konsumsi dan perilaku belanja. Pasar daring seperti Shopee, Tokopedia, dan Bukalapak beroperasi sebagai ekosistem bisnis yang menghubungkan pembeli dan penjual dalam skala besar, menawarkan jutaan pilihan produk, beragam metode pembayaran, dan strategi promosi yang menarik untuk meningkatkan daya beli masyarakat.

*e-commerce* di Indonesia berkembang sangat pesat, akan tetapi penjual masih kesulitan untuk mengidentifikasi *marketplace* terbaik untuk menjual produk yang sesuai. Tantangan ini biasanya disebabkan oleh kurangnya pengetahuan yang komprehensif tentang potensi masing-masing *marketplace* serta keterbatasan kemampuan analisis data. Akibatnya, banyak keputusan bisnis yang dibuat berdasarkan spekulasi. Oleh karena itu, untuk membantu penjual mengetahui potensi pasar pada *marketplace* seperti Shopee, Tokopedia, dan Bukalapak, diperlukan pendekatan analitis.

Berdasarkan laporan pada tahun 2022 dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menyatakan bahwa dengan tingkat penetrasi 77,02%, saat ini terdapat 210.026.769 pengguna internet di Indonesia dari total populasi 272.682.600 jiwa [1]. Angka ini menggambarkan potensi ekonomi digital Indonesia yang sangat besar, yang terus berkembang setiap tahunnya. Aktivitas transaksi daring di *marketplace* secara langsung dipengaruhi oleh peningkatan pengguna internet. Selain memperluas jangkauan pemasaran, fenomena ini juga meningkatkan persaingan bagi perusahaan yang bersaing untuk mendapatkan perhatian pelanggan.

*Volume* transaksi yang tinggi tidak selalu berarti pemilik bisnis akan menghasilkan keuntungan terbesar. Potensi pasar yang besar merupakan tugas tersulit untuk memaksimalkan harga agar menemukan produk yang sesuai, taktik pemasaran, dan perencanaan stok. Kehabisan stok dapat mengakibatkan hilangnya peluang penjualan dan penurunan kepuasan pelanggan, produk yang tidak terjual atau permintaan yang salah perhitungan dapat menyebabkan kelebihan stok, yang mengakibatkan biaya penyimpanan dan modal menganggur. Untuk memperkirakan pola penjualan dan mengklasifikasikan produk berdasarkan potensi pasarnya, skenario ini memerlukan strategi analitis berbasis data [2].

Salah satu sumber informasi yang berguna dalam hal ini adalah pengelolaan data penjualan yang dapat diakses di *platform marketplace* [3]. Variabel penting termasuk harga produk, *volume* penjualan, ulasan pelanggan, peringkat, dan tingkat diskon semuanya termasuk dalam data ini [4][5]. Bisnis dapat menentukan tren konsumsi pelanggan dan memperkirakan potensi pasar setiap produk dengan memanfaatkan metodologi analisis dan kategorisasi yang sesuai. *K-Nearest Neighbor* (KNN), algoritma pembelajaran mesin yang populer untuk tujuan ini, mengelompokkan data menurut seberapa dekat atribut antar objek dalam ruang fitur [2].

Manajer toko dan pemilik bisnis dapat mengklasifikasikan produk dari yang terlaris sampai dengan yang tidak laris sebagai salah satu cara strategis untuk menganalisis potensi pasar agar dapat membantu dalam membuat keputusan yang tepat. Produk menawarkan wawasan penting tentang tren dan perilaku konsumen selain mencerminkan permintaan dan preferensi pasar saat ini [6]. Penelitian sebelumnya telah menggunakan berbagai teknik pembelajaran mesin, seperti *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Decision Tree*. Namun, karena sifatnya yang mudah namun efisien dalam mengidentifikasi pola berdasarkan kedekatan data dalam ruang fitur, metode *K-Nearest Neighbor* (KNN) terus menjadi salah satu algoritma yang paling banyak digunakan dalam skenario klasifikasi penjualan. Sebagai teknik pembelajaran terawasi, *K-Nearest Neighbor* (KNN) dapat menggabungkan potensi pasar suatu produk di beberapa *platform marketplace* dengan mengklasifikasikan data produk berdasarkan kesamaan fitur [7].

Efisiensi KNN dalam analisis sentimen dan klasifikasi produk di industri *e-commerce* telah ditunjukkan dalam sejumlah studi sebelumnya. Selain itu, studi lain menunjukkan bahwa penggunaan KNN untuk menganalisis evaluasi aplikasi Shopee di *Google Playstore* dapat mencapai akurasi 70%, dengan kemungkinan peningkatan lebih lanjut melalui optimasi teknik pengambilan sampel [8]. Hiperparameter penting yang berdampak langsung pada kinerja model KNN adalah pemilihan rasio pemisahan data dan penyetelan nilai K. Skenario dengan pemisahan 70:30 dan nilai K 9 berkinerja terbaik, menurut pengujian eksplisit studi ini terhadap beberapa rasio pemisahan dan nilai K [9]. Hal ini menunjukkan bahwa peningkatan akurasi dan ketergantungan model dalam aplikasi *e-commerce* memerlukan penyesuaian hiperparameter, yang bukan merupakan langkah sederhana. Selain itu, kinerja khusus kelas terutama *recall* untuk kelas minoritas sangat dipengaruhi oleh ketidakseimbangan data [9]. Lebih dari sekadar skor prediksi,

interpretasinya (berdasarkan "tetangga terdekat") mungkin menawarkan wawasan yang berguna untuk manajemen inventaris atau pemasaran yang ditargetkan. Salah satu contoh jelas dari informasi yang dapat ditindaklanjuti adalah tren pembelian berbasis gender yang ditemukan[10].

Kategori penjualan di industri, termasuk elektronik, barang digital dan kebutuhan sehari-hari, KNN dapat memprediksi penjualan produk dengan tingkat akurasi yang tinggi. Di sejumlah *marketplace*, parameter  $k$  yang benar dapat meningkatkan akurasi kategorisasi potensi pasar produk[2]. KNN mencapai akurasi 83,29% dalam kategorisasi ulasan, menurut studi berbeda yang membandingkannya dengan *Naïve Bayes* untuk analisis sentimen pada Shopee dan TikTok Shop [11]. A. Oktian Permana selanjutnya memverifikasi bahwa KNN masih berlaku karena kemampuan adaptasinya terhadap keadaan data yang berbeda, meskipun *Naïve Bayes* memiliki akurasi yang lebih besar dalam studi kasus Shopee [12]. Penelitian ini mengantisipasi bahwa dengan menggabungkan jenis analisis ini dengan data produk dari Shopee, Tokopedia, dan Bukalapak, akan menciptakan model klasifikasi yang lebih akurat meramalkan potensi pasar, yang dapat digunakan pelaku bisnis untuk membuat rencana pemasaran dan mengelola inventaris secara tepat.

Berdasarkan latar belakang tersebut, teknik analisis utama yang digunakan dalam penelitian ini adalah algoritma *K-Nearest Neighbors* (KNN), yang bertujuan untuk mengevaluasi potensi pasar pada *platform marketplace* dalam hal *volume* penjualan. Studi ini berfokus pada Shopee, Tokopedia, dan Bukalapak, tiga *platform* utama yang pada tahun tersebut menguasai lingkungan perdagangan digital Indonesia. Ketiga *platform* ini menawarkan gambaran representatif tentang bisnis *e-commerce* nasional karena dipilih berdasarkan popularitas, pangsa pasar, dan jangkauan teknik pemasarannya.

Penelitian ini berfokus pada pembagian data produk ke dalam dua kategori utama yaitu produk terlaris dan tidak dengan mempertimbangkan berbagai kriteria kunci yang diyakini memengaruhi tingkat penjualan. Variabel-variabel ini meliputi harga produk, *volume* penjualan, jumlah diskon, penilaian konsumen, dan jumlah ulasan yang diterima. Untuk memastikan bagaimana setiap variabel memengaruhi kemungkinan suatu produk menjadi produk terlaris di pasar, variabel tersebut akan diukur dan dikaji. Analisis ini diharapkan dapat memberikan saran dan rekomendasi bagi berbagai pelaku usaha untuk meningkatkan daya saing dan memaksimalkan keuntungan dalam ekonomi digital yang sangat kompetitif.

## 2. Metode Penelitian

Menjelaskan kronologis penelitian, termasuk desain penelitian, prosedur penelitian (dalam bentuk algoritma, *pseudocode*, atau lainnya), bagaimana menguji dan mengumpulkan data [1], [3]. Deskripsi alur penelitian harus didukung oleh referensi sehingga penjelasan yang ditulis dapat diterima secara ilmiah

### 2.1. Data Penelitian

Dalam penelitian ini, dilakukan analisis performa model *K-Nearest Neighbors* untuk melakukan klasifikasi data produk menjadi kategori laris dan tidak. Data yang digunakan diperoleh dengan cara scraping menggunakan aplikasi *BigSeller* dilakukan secara langsung melalui website *marketplace* shopee (<https://shopee.co.id/>), bukalapak (<https://www.bukalapak.com/>) dan tokopedia (<https://www.tokopedia.com/>) pada bulan maret tahun 2022.

Dataset awal terdiri dari 2.250 record yang mencakup berbagai parameter untuk analisis potensi pasar *marketplace* yaitu kategori barang, *marketplace*, nama barang, harga, jumlah ulasan, rating toko, jumlah terjual, diskon serta kategori penjualan.

Tabel 1. Ilustrasi data penjualan

No	Kategori Barang	Marketplace	Harga	Jumlah Ulasan	Rating Toko	Terjual	Diskon	Kategori penjualan
1	Fashion	Shopee	160,000	3	5.0	10	0	Rendah
2	Fashion	Shopee	18,000	50	5.0	376	0	Tinggi
3	Fashion	Shopee	19,150	19000	4.2	63000	65%	Tinggi
4	Fashion	Shopee	19,500	196	4.8	574	0	Tinggi
5	Fashion	Shopee	7,000	183	4.9	10000	0	Tinggi
...	...	...	...	...	...	...	...	...
2247	Otomotif	Bukalapak	300,000	0	0	0	0	Rendah
2248	Otomotif	Bukalapak	45,000	1	5.0	1	10%	Rendah
2249	Otomotif	Bukalapak	67,500	0	0	1	10%	Rendah
2250	Otomotif	Bukalapak	58,500	10	4.9	15	10%	Rendah

## 2.2. Metode Penelitian

Dalam penelitian ini menggunakan metode Machine Learning yaitu K-Nearest Neighbors untuk melakukan klasifikasi potensi penjualan menjadi tinggi sedang dan rendah.

### 2.2.1. K-Nearest Neighbors

Pendekatan pembelajaran berbasis instance-based learning untuk regresi atau klasifikasi adalah algoritma K-Nearest Neighbors (KNN). Kelompok teknik klasifikasi mencakup sejumlah algoritma pembelajaran mesin, termasuk KNN[13]. Menemukan jarak terkecil antara data baru dan data dalam dataset yang kelasnya diketahui adalah ide dasar di balik KNN. Karena manfaatnya dalam mengelola data non-linier dan kemudahan penggunaannya, pendekatan ini sering digunakan dalam analisis prediktif[14]. Menemukan nilai K adalah langkah pertama dalam algoritma K-Nearest Neighbor. Langkah selanjutnya adalah menghitung kedekatan nilai antara objek setelah nilai K ditetapkan[15]. Objek yang dipilih untuk memenuhi nilai K dalam K-Nearest Neighbor ditentukan oleh jarak Euclidean. Menemukan jarak terdekat antara objek yang baru ditambahkan dan objek yang disimpan sebelumnya dalam data dikenal sebagai jarak Euclidean [16].

Jarak antara dua titik pada data training dan titik pada data testing dapat didefinisikan dengan rumus Euclidean, seperti berikut.

$$d = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (1)$$

Dengan

$d$  : jarak *Euclidean*

$x_{2i}$  : nilai pada data testing ke -  $i$

$x_{1i}$  : nilai pada data training ke -  $i$

$p$  : banyaknya atribut

Rumus ini digunakan untuk menunjukkan jarak antara dua titik data, dengan  $x_{1i}$  adalah koordinat fitur titik pertama dan  $x_{2i}$  adalah koordinat fitur titik kedua. Dengan menggunakan selisih nilai fitur setiap titik, rumus ini menghitung jarak antara dua titik dalam ruang berdimensi -  $p$  [17].

### 2.2.2. Evaluasi model

Dalam penelitian ini, kinerja klasifikasi diukur menggunakan confusion matrix untuk menguji metodologi. confusion matrix: true positives, true negatives, false positives, dan false negatives merupakan teknik evaluasi yang mencirikan hasil proses klasifikasi [18].

1. Accuracy, proporsi data yang diklasifikasikan dengan tepat oleh sistem menggunakan rumus di bawah ini.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

2. Precision, presentase akurasi antara data yang diminta dengan hasil klasifikasi menggunakan rumus berikut.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

3. Recall, keberhasilan model klasifikasi dalam menentukan kembali informasi atau hasil nilai dengan menggunakan rumus berikut.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

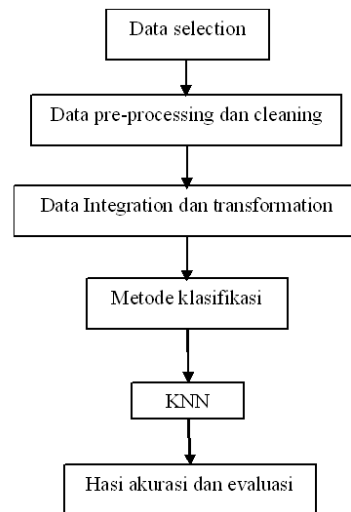
4. F1-Score, suatu perbandingan nilai rata-rata dari precision dan recall dengan menggunakan rumus berikut.

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (5)$$

Menggabungkan pengukuran-pengukuran ini menjadi metrik komposit, seperti skor F1, merupakan strategi yang lebih unggul untuk menangani kumpulan data yang terdistribusi tidak merata [19]. Parameter kinerja termasuk presisi, perolehan kembali, dan skor F1 kemudian dihitung secara bersamaan dalam klasifikasi multikelas menggunakan rata-rata mikro.

### 2.2.3. Alur penelitian

Adapun tahapan merancang dan mengimplementasikan sistem dimana sistem dirancang dan diimplementasikan sesuai dengan tujuan penelitian yang telah difokuskan.



Gambar 1. Alur Analisis KNN

Data penjualan digunakan sebagai dataset awal penelitian ini. Prosedur seleksi digunakan untuk mengidentifikasi karakteristik yang diperlukan untuk klasifikasi data penjualan dari dataset asli. Setelah itu, pre-processing dan cleaning dilakukan untuk menghilangkan data kosong (missing values dan noisy) serta nilai kosong dari atribut. Integrasi data, yang melibatkan penggabungan semua variabel dataset ke dalam satu tabel, dilakukan setelah tahap pre-processing data. Selain itu, menghindari abnormalitas merupakan tujuan dari prosedur normalisasi data. Algoritma KNN, yang mengklasifikasikan objek berdasarkan data pembelajaran yang paling dekat dengan objek tersebut, digunakan untuk proses klasifikasi setelah transformasi selesai.

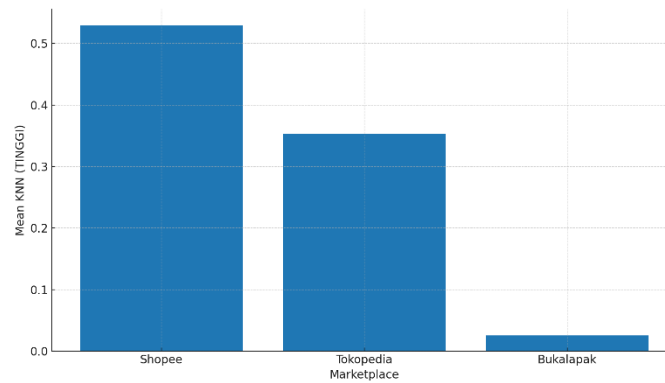
### 3. Hasil dan Pembahasan

Pada pembahasan dilakukan evaluasi dan analisa hasil klasifikasi dari metode yaitu KNN terhadap data tiga marketplace yaitu shopee bukalapak dan tokopedia yang diperoleh dari website resmi <https://www.shopee.co.id/>, <https://www.bukalapak.com/>, <https://www.tokopedia.com/> pada bulan maret tahun 2022.

Tabel 1. Analisis menggunakan KNN

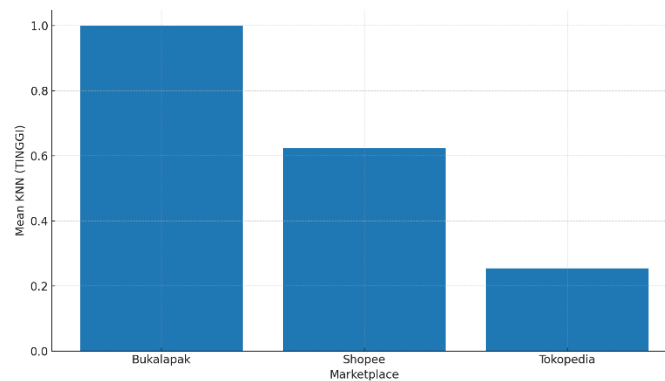
No	Kategori Penjualan	Kategori Barang	Knn Prediksi	Knn (Rendah)	Knn (Sedang)	Knn (Tinggi)
1	Tinggi	Fashion	Tinggi	0	0	1
2	Tinggi	Fashion	Tinggi	0	0	1
3	Tinggi	Fashion	Rendah	0.666667	0	0.333333
4	Tinggi	Fashion	Tinggi	0	0	1
...	...	...	...	...	...	...
2249	Rendah	Otomotif	Rendah	1	0	0
2250	Rendah	Otomotif	Rendah	1	0	0

Berdasarkan hasil analisis data menggunakan KNN, terlihat pola yang cukup jelas terkait performa penjualan di tiga kategori utama Fashion, Elektronik, dan Otomotif. Pada tiga marketplace besar Shopee, Bukalapak, dan Tokopedia.



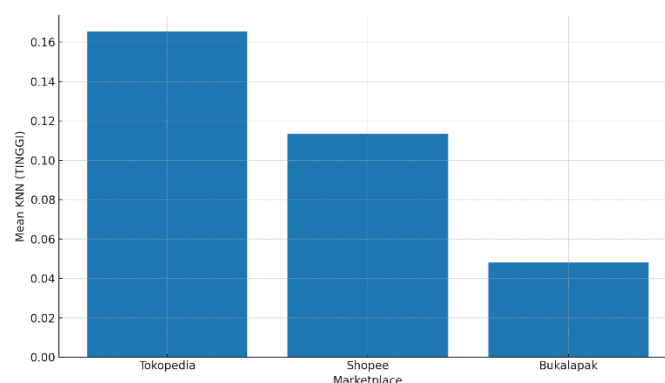
Gambar 2. KNN kategori Fashion

Dengan skor KNN rata-rata (TINGGI) sebesar 0,529, Shopee berada di posisi pertama untuk kategori fashion. Ukuran sampel yang memadai dan distribusi label yang diproyeksikan menunjukkan dominasi kelas "Tinggi" memperkuat kesimpulan model KNN bahwa mayoritas produk fashion di shopee memiliki kemungkinan penjualan yang tinggi, seperti yang ditunjukkan pada gambar ini. Hal ini menegaskan bahwa hasil model sejalan dengan realitas pasar, karena shopee dikenal kuat di bidang fashion dan gaya hidup.



Gambar 3. KNN kategori Elektronik

Bukalapak memegang posisi teratas dalam kategori elektronik dengan skor KNN (tinggi) rata - rata 1.000. Dengan nilai ideal ini, diantisipasi bahwa sampel lengkap perangkat elektronik di Bukalapak yang termasuk dalam dataset ini akan memiliki potensi tinggi untuk penjualan. Meskipun hasil ini tampak luar biasa, penting untuk memperhitungkan ukuran sampel. Jika kumpulan data tidak terlalu besar, nilai ideal ini harus dievaluasi dengan hati -hati. Namun, secara umum, dominasi ini dapat dibenarkan oleh elemen-elemen seperti reputasi penjual, reputasi ulasan, dan harga kompetitif untuk kategori elektronik di Bukalapak.



Gambar 4. KNN kategori Otomotif

Sebaliknya, kategori otomotif menunjukkan tren yang berbeda. Dengan rata-rata skor (tinggi) 0,165, Tokopedia ada di teratas dan secara signifikan lebih rendah dari kategori lainnya. Ini menunjukkan bahwa, sementara Tokopedia terus mengungguli Shopee dan Bukalapak, potensi keseluruhan untuk penjualan tinggi dalam kategori otomotif dalam dataset ini agak rendah. Perbedaan ini mungkin disebabkan oleh fitur-fitur pasar otomotif, yang mencakup siklus pembelian yang lebih lama, nilai transaksi tinggi, dan keputusan untuk membeli yang memperhitungkan rincian yang lebih teknis daripada pertimbangan promosi yang cepat. Peringkat teratas dan terbawah memiliki margin perbedaan yang berbeda jika dibandingkan dengan total. Sebaliknya, industri otomotif memiliki perbedaan tipis dan skor total yang rendah, menunjukkan persaingan yang lebih merata tetapi dengan potensi pasar yang lebih kecil. Dalam Fashion dan elektronik, kesenjangannya sangat besar, membuat pemenangnya lebih unggul. Oleh karena itu, tindakan terbaik adalah memanfaatkan potensi fashion untuk Shopee, melestarikan dominasi Bukalapak dalam elektronik, dan berkonsentrasi pada pendukung daftar Tokopedia di sektor otomotif sambil mencari metode untuk meningkatkan indikator penjualan yang tinggi di daerah itu.

Tabel 2. Hasil pada Model KNN dengan Menggunakan k-fold cross validation

Model	Metric	Score
KNN	AUC	0.918
	CA	0.846
	F1	0.842
	Prec	0.841
	Recall	0.846
	MCC	0.730

Berdasarkan hasil evaluasi kinerja model KNN, yang digunakan untuk menganalisis potensi penjualan di berbagai marketplace, bekerja sangat baik sebagai alat bantu pengambilan keputusan strategis. Kemampuan model ini dalam membedakan produk dengan potensi penjualan tinggi dan rendah ditunjukkan oleh nilai AUC sebesar 0,9185. Dalam analisis marketplace, ini berarti bahwa meskipun atribut produk di berbagai kategori berbeda secara signifikan, model ini dapat dengan andal menentukan produk mana yang kemungkinan besar akan sukses di pasar dan mana yang seringkali berkinerja buruk. Dengan kata lain, model ini memiliki "insting" yang kuat untuk memilah produk-produk potensial lebih sering di Shopee, Bukalapak, dan Tokopedia, dengan akurasi yang hampir sempurna pada skala evaluasi ROC.

Mayoritas prakiraan menggambarkan situasi aktual di lapangan dengan tepat, berdasarkan akurasi klasifikasi (CA) sebesar 84,62%. Hal ini penting dalam konteks bisnis karena ketidakakuratan prakiraan dapat memengaruhi pilihan strategis seperti memilih marketplace, mengalokasikan dana untuk promosi, atau memprioritaskan stok. Akurasi model dalam memilih produk berpotensi tinggi (minim false positive) dan kemampuannya untuk menangkap mayoritas produk potensial di pasar (minim false negative) dikonfirmasi oleh keseimbangan antara Precision (84,12%) dan recall (84,62%) serta F1 Score sebesar 84,22%. Hal ini memastikan bahwa peluang pada produk yang tepat, tidak hilang sekaligus mengurangi risiko mengalokasikan sumber daya untuk produk yang salah.

Gambaran menyeluruh tentang stabilitas kinerja model di semua kelas ditunjukkan oleh MCC (Matthews Correlation Coefficient) yang sebesar 0,7302. Karena data terkadang tidak seimbang misalnya, produk dengan penjualan tinggi biasanya jauh lebih sedikit daripada produk yang stagnan hal ini sangat penting di pasar. Meskipun distribusinya tidak seimbang, MCC yang tinggi menunjukkan bahwa model tersebut mempertahankan kualitas prediktif, menjamin bahwa temuannya dapat diterapkan dan dapat dipercaya untuk pengambilan keputusan di berbagai industri, termasuk mode, elektronik, dan otomotif.

Secara praktis, model kNN ini dapat berfungsi sebagai bagian fundamental dari analisis potensi penjualan dan sistem rekomendasi. Model ini dapat digunakan oleh marketplace atau penjual untuk menentukan produk mana yang paling sesuai untuk promosi, menetapkan kebijakan harga, atau bahkan membuat paket penawaran berdasarkan peluang pembelian terbaik. Misalnya, model ini dapat digunakan untuk menentukan kombinasi produk dan taktik diskon paling sukses dalam industri fashion yang kompetitif. Meskipun kategori otomotif memiliki siklus pembelian yang lebih panjang, model ini dapat membantu menargetkan segmen pasar yang tepat dengan iklan yang efektif. Dalam kategori elektronik, model ini dapat memberikan justifikasi atas penilaian terhadap ketersediaan barang-barang mewah atau gawai terbaru. Karena kinerjanya yang kuat, model KNN dapat digunakan sebagai alat analisis data, tetapi juga sebagai pemandu strategi bisnis di dunia marketplace. Namun, evaluasi berkala tetap diperlukan, seperti halnya model prediksi lainnya. Pembaruan model secara berkala sangat penting untuk

mempertahankan tingkat akurasi dan relevansi yang tinggi karena pola data dapat dipengaruhi oleh perubahan perilaku pelanggan, tren industri, atau peluncuran produk baru.

#### 4. Kesimpulan

Model KNN menunjukkan performa yang sangat baik dan konsisten. Model ini dapat secara akurat membedakan produk dengan potensi penjualan tinggi sedang dan rendah, berdasarkan nilai AUC sebesar 0,9185. Nilai akurasi (CA) sebesar 84,62% mendukung temuan bahwa mayoritas prakiraan sesuai dengan kondisi aktual di lapangan. Model ini dapat mendeteksi produk target dan jarang melewatkan produk prospektif, terbukti dari keseimbangan antara *Precision* (84,12%) dan *Recall* (84,62%), yang menghasilkan *F1 Score* sebesar 84,22%. Di sisi lain, nilai MCC sebesar 0,7302 menunjukkan bahwa performa model tetap stabil meskipun menggunakan data yang sering kali menunjukkan ketidakseimbangan jumlah pproduk dengan penjualan tertinggi sedang dan rendah. Pola potensi penjualan yang jelas dapat dilihat ketika memeriksa hasil KNN berdasarkan *marketplace* dan kategori penjualan. *Shopee* menerima skor tertinggi (0,5293) pada kategori *fashion*, Bukalapak menerima skor sempurna 1,0000 dalam kategori elektronik dan Tokopedia memiliki skor tertinggi dalam kategori otomotif (0,1653). Pemilik produk dapat menggunakan pemetaan ini untuk membuat strategi pemasaran yang lebih terfokus, seperti berkonsentrasi pada *fashion* di *Shopee*, elektronik di Bukalapak, dan otomotif di Tokopedia. Ini akan membantu mereka mengoptimalkan promosi penjualan, manajemen stok, dan alokasi anggaran promosi untuk memanfaatkan peluang pasar sebaik mungkin. Membandingkan KNN dengan algoritma lain, seperti *C45*, *Random Forest* atau *XGBoost*, disarankan untuk studi selanjutnya guna menilai potensi peningkatan akurasi. Kualitas prediksi juga dapat ditingkatkan dengan menggunakan data eksternal, seperti tren pencarian dan ulasan pelanggan. Taktik bisnis yang lebih terarah juga dapat ditingkatkan dengan menciptakan sistem pendukung keputusan dan memeriksa segmentasi pasar berdasarkan preferensi konsumen dan lokasi geografis.

#### Daftar Pustaka

- [1] Apjii, "Asosiasi Penyelenggara Jasa Internet Indonesia (Apjii) ," Jun. 2022, Accessed: Aug. 13, 2025. [Online]. Available: [https://Apjii.Or.Id/Berita/D/Apjii-Di-Indonesia-Digital-Outlook-2022\\_857](https://Apjii.Or.Id/Berita/D/Apjii-Di-Indonesia-Digital-Outlook-2022_857)
- [2] Roni S And C. Crysdiان, "Jurnal Teknologi Dan Manajemen Informatika Studi Literature Analisis Potensi Pasar Marketplace Terhadap Penjualan Article Info Abstract," *Jurnal Teknologi Dan Manajemen Informatika (Jtmi)*, Vol. 8, No. 2, Pp. 134–142, 2022, [Online]. Available: [Http://Http://Jurnal.Unmer.Ac.Id/Index.Php/Jtmi](http://Http://Jurnal.Unmer.Ac.Id/Index.Php/Jtmi)
- [3] C. Y. Hazizah And T. Widiyaningtyas, "Analisis Metode Collaborative Filtering Menggunakan Knn Dan Svd++ Untuk Rekomendasi Produk E-Commerce Tokopedia," *Edumatic: Jurnal Pendidikan Informatika*, Vol. 8, No. 2, Pp. 595–604, Dec. 2024, Doi: 10.29408/Edumatic.V8i2.27793.
- [4] R. Rahmawati And E. Suwarni, "Ulasan Produk Dan Jumlah Produk Terjual Dampaknya Pada Keputusan Pembelian Di Marketplace Shopee," *Jurnal Dimensi* /, Vol. 03 No 1, No. 1, Pp. 46–53, Feb. 2023, Doi: <https://doi.org/10.32897/Dimensi.V3i1.2539>.
- [5] S. Bahri And A. M. Widodo, "Penerapan Algoritma Pengklasifikasi Untuk Mengukur Kepuasan Pelanggan E-Commerce (Studi Kasus : Shopee)," *Marketgram Journal*, Vol. 03, No. 01, Pp. 101–114, 2024, [Online]. Available: <https://E-Journal.Naurendigiton.Com/Index.Php/Mj>
- [6] Intan Purwasih, Kiki Setiawan, Francis Matheos Sarimole, And Tundo, "Klasifikasi Penjualan Produk Terlaris Pada Kedai Ira Dengan Menggunakan Algoritma Naïve Bayes Dan Algoritma K-Nearest Neighbor," *Ijccs*, Vol. 8, No.2, Pp. 695–706, Aug. 2024, Accessed: Aug. 13, 2025. [Online]. Available: <https://Jurnal.Polsri.Ac.Id/Index.Php/Teknika/Article/View/9100>
- [7] R. S. Nurhalizah, R. Ardianto, And P. Purwono, "Analisis Supervised Dan Unsupervised Learning Pada Machine Learning: Systematic Literature Review," *Jurnal Ilmu Komputer Dan Informatika*, Vol. 4, No. 1, Pp. 61–72, Aug. 2024, Doi: 10.54082/Jiki.168.
- [8] T. Fardaningsih And A. Lutfiyani, "Insologi: Jurnal Sains Dan Teknologi Perbandingan Analisis Sentimen Pengguna Aplikasi Shopee Dan Lazada Pada Situs Google Play Store Menggunakan Algoritma K-Nearst Neighbor Dan Naive Bayes," *Media Cetak*, Vol. 4, No. 3, Pp. 563–578, 2025, Doi: 10.55123/Insologi.V4i3.5646.
- [9] M. Saifurridho And U. Hayati, "Jurnal Informatika Terpadu Analisis Algoritma K-Nearest Neighbor Terhadap Sentimen Pengguna Aplikasi Shopee," *Jurnal Informatika Terpadu*, Vol. 10, No. 1, Pp. 21–26, 2024, [Online]. Available: <https://Journal.Nurulfikri.Ac.Id/Index.Php/Jit>



- [10] A. Verma, C. Nagar, S. Haryani, And S. Jain, "Prediction Of E-Commerce Shoppers' Purchasing Intention Using Knn Algorithm," In *Atlantis Press*, 2025, Pp. 63–72. Doi: 10.2991/978-94-6463-716-8\_6.
- [11] S. Chairunnisa Nandaresta And C. Warman, "Sismatik (Seminar Nasional Sistem Informasi Dan Manajemen Informatika) Universitas Nusa Putra, 12 Agustus 2023," *Sismatik (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, Vol. 3, 2023, Accessed: Aug. 13, 2025. [Online]. Available: <https://Sismatik.Nusaputra.Ac.Id/Index.Php/Sismatik/Article/View/216/131>
- [12] A. Oktian Permana And Sudin Saepudin, "Perbandingan Algoritma K-Nearst Neighbor Dan Naïve Bayes Pada Aplikasi Shopee," *Jurnal Coscitech (Computer Science And Information Technology)*, Vol. 4, No. 1, Pp. 25–32, Apr. 2023, Doi: 10.37859/Coscitech.V4i1.4474.
- [13] D. Sugiarto *Et AL.*, "Perbandingan Kinerja Klasifikasi Sentimen Ulasan Produk Pembelian Beras Di Marketplace Shopee," *Jurnal Teknologi Informasi : Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika*, Vol. 17, No. 1, 2023, Doi: 10.47111/Jti.
- [14] Ahmad Muflih Wafir And Zaehol Fatah, "Prediksi Produk Penjualan Di Supermarket Dengan Metode Algoritma K-Nearest Neighbors (Knn)," *Jurnal Ilmiah Sains Teknologi Dan Informasi*, Vol. 3, No. 1, Pp. 75–82, Dec. 2024, Doi: 10.59024/Jiti.V3i1.1056.
- [15] A. Halim Lubis And R. Rizky Nasution, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Elektronik Terlaris Menggunakan Metode K-Nearest Neighbor," 2024. [Online]. Available: [Http://Jurnal.Goretanpena.Com/Index.Php/Jssr](http://Jurnal.Goretanpena.Com/Index.Php/Jssr)
- [16] B. Susilo, N. A. Ramdhan, And O. S. Bachri, "Penerapan Algoritma K-Nearest Neighbor Untuk Prediksi Penjualan Produk Digital," *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, Vol. 4, No. 4, Pp. 1466–1476, Sep. 2024, Doi: 10.57152/Malcom.V4i4.1517.
- [17] N. Nabiilah, S. Rohimah, And S. K. P. Loka, "Analyzing Customer Sentiment Towards Marketplace Reviews Using Classification Algorithms," *Ijatis: Indonesian Journal Of Applied Technology And Innovation Science*, Vol. 2, No. 1, Pp. 23–30, Feb. 2025, Doi: 10.57152/Ijatis.V2i1.1774.
- [18] A. K. Febrian, Y. H. Chrisnanto, D. Pupita, N. Sabrina, And J. Achmad Yani, "Snestik Seminar Nasional Teknik Elektro, Sistem Informasi, Dan Teknik Informatika Studi Komparasi Metode Klasifikasi K-Nearest Neighbor Dan Naïve Bayes Dalam Mengidentifikasi Kepuasan Pelanggan Terhadap Produk," *Snestik*, Pp. 333–338, Mar. 2022, Doi: 10.31284/P.Snestik.2022.2717.
- [19] F. S. Aditama, D. Krismawati, And S. Pramana, "Multiclass Classification Of Marketplace Products With Machine Learning," *Media Statistika*, Vol. 17, No. 1, Pp. 25–35, Oct. 2024, Doi: 10.14710/Medstat.17.1.25-35.